



Towards cost-effective side-chain isotope labelling of proteins expressed in human cells

Martina Rosati¹ · Letizia Barbieri^{1,2} · Matus Hlavac³ · Sarah Kratzwald³ · Roman J. Lichtecker^{3,4} · Robert Konrat^{3,5} · Enrico Luchinat^{1,2,6} · Lucia Banci^{1,2,6}

Received: 23 February 2024 / Accepted: 17 July 2024
© The Author(s) 2024

Abstract

Side chain isotope labelling is a powerful tool to study protein structure and interactions by NMR spectroscopy. ¹H, ¹³C labelling of side-chain methyl groups in a deuterated background allows studying large molecules, while side-chain aromatic groups are highly sensitive to the interaction with ligands, drugs, and other proteins. In *E. coli*, side chain labelling is performed by substituting amino acids with isotope-labelled precursors. However, proteins that can only be produced in mammalian cells require expensive isotope-labelled amino acids. Here we provide a simple and cost-effective method to label side chains in mammalian cells, which exploits the reversible reaction catalyzed by endogenous transaminases to convert isotope-labelled α -ketoacid precursors. We show by in-cell and in-lysate NMR spectroscopy that replacing an amino acid in the medium with its cognate precursor is sufficient to achieve selective labelling without scrambling, and how this approach allows monitoring conformational changes such as those arising from ligand binding.

Keywords Protein labelling · Alpha-ketoacid · In-cell NMR · HEK293T cells · Mammalian expression · Transaminases

Introduction

Interactions underpin all biological processes. They occur between proteins, between proteins and their biological ligands and inhibitors, or between proteins and synthetic molecules. NMR spectroscopy is ideally suited to investigate such interactions at atomic resolution, in physiologically

relevant settings, e.g. in intact cells and lysates (Theillet and Luchinat 2022). The various types of interactions involve specific types of amino acids. Therefore, it is essential to develop novel, efficient, and cost-effective technologies to examine interactions by looking at the side chains of the specific amino acids relevant for molecular interactions (Schörghuber et al. 2018). Furthermore, structural characterization of large proteins by NMR requires residue-specific labelling in order to improve spectral resolution and sensitivity, and to reduce signal overlap. Two side chain labelling types are particularly useful to investigate protein structure, dynamics and interactions: ¹H, ¹³C labelling of aromatic residues and methyl-¹H, ¹³C labelling of aliphatic residues. Aromatic residues are often involved in the interaction with other proteins, ligands or drugs, and take part in enzymatic activities (Serrano et al. 1991; Schörghuber et al. 2018; Ninković et al. 2020). On the other hand, labelled methyl groups of aliphatic residues are highly useful when investigating slow-tumbling macromolecules, thanks to their high intrinsic mobility and three-fold proton multiplicity, which result in a greatly enhanced sensitivity. Methyl labelling allows for the recording of NMR experiments of large proteins that would otherwise feature exceedingly broadened signals due to fast transverse relaxation (Kerfah et al. 2015). In cells, a similar

✉ Enrico Luchinat
eluchinat@cerm.unifi.it

✉ Lucia Banci
banci@cerm.unifi.it

¹ CERM — Magnetic Resonance Center, Università degli Studi di Firenze, Sesto Fiorentino, Italy

² Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine — CIRMMP, Sesto Fiorentino, Italy

³ MAG-LAB GmbH, Vienna, Austria

⁴ Institute of Organic Chemistry, University of Vienna, Vienna, Austria

⁵ Department of Structural and Computational Biology, Max Perutz Laboratories, University of Vienna, Vienna, Austria

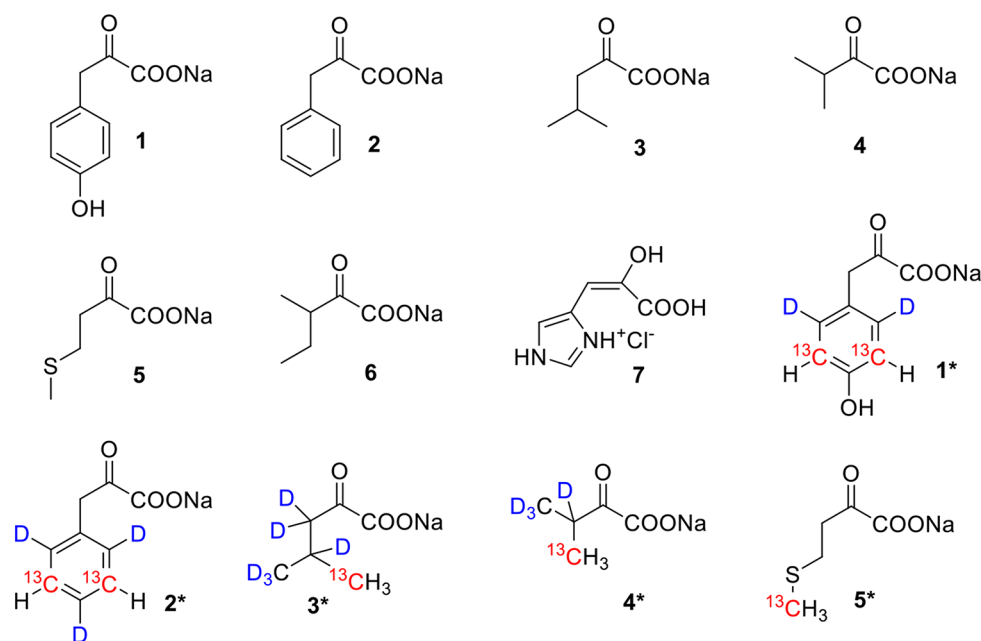
⁶ Dipartimento di Chimica, Università degli Studi di Firenze, Sesto Fiorentino, Italy

scenario is also encountered when studying small globular proteins which interact with large, abundant cellular components (Barbieri et al. 2015; Burz et al. 2019). In both cases, the presence of a deuterated side chain background further increases the sensitivity by decreasing the ^1H spin relaxation rates. Moreover, sparse labelling schemes greatly simplify the spectra of large proteins by reducing spectral overlap.

In *E. coli*, side chain labelling is typically performed by substituting amino acids with the cognate labelled α -ketoacid precursors to the growth medium, as these compounds can be synthesized from commercially available isotope sources without the need of creating chiral carbons by elaborate organic synthesis. Isotope-labelled α -ketoacids are then effectively incorporated into the host organism's biosynthetic pathways and converted into the respective amino acids by bacterial transaminases, which substitute the keto group with an amino group in vivo (Goto et al. 1999; Lichtenecker et al. 2004, 2013a, c). Although this method is extremely effective, it has never been applied to those proteins, which require production in mammalian cells. Indeed, previous reports on selective isotope labelling in mammalian cells had to make use of synthetically more demanding and thus more expensive isotope-labelled amino acids as additives to substitute the unlabelled target residues (Werner et al. 2008; Banci et al. 2013). In mammals, transaminases reversibly catalyze the first catabolic reaction in the degradation of amino acids, transferring their amino group to a ketoacid substrate, usually α -ketoglutarate, in order to obtain

the corresponding α -ketoacid and glutamate (Caligiore et al. 2022). As some transaminases are known to be expressed inside mammalian tissues (Caligiore et al. 2022; Neinst et al. 2019; Toyokawa et al. 2021; Sivaraman and Kirsch 2006; Mehere et al. 2010; Cooper 2004), we reasoned that supplementing expression media with α -ketoacids should result in the biosynthesis of the corresponding amino acids in cultured mammalian cells, thus allowing the expression of proteins with amino acid-specific labelling schemes at a lower cost.

In this work, we show that HEK293T cells supplied with custom-made culture media containing an α -ketoacid precursor in place of the corresponding amino acid result in a high-yield transient expression of the desired proteins. These experiments were performed with the unlabelled α -ketoacid precursors of tyrosine (p-hydroxyphenylpyruvate **1**), phenylalanine (phenylpyruvate **2**), leucine (α -ketoisocaproate **3**), valine (α -ketoisovalerate **4**), methionine (4-methylthio- α -ketobutanoic acid **5**), isoleucine (3-methyl- α -ketovalerate **6**) and histidine (4-(2-carboxy-2-hydroxyvinyl)-*1H*-imidazolium chloride **7**). Selective side-chain labelling protocols were demonstrated on two model proteins, carbonic anhydrase II (CA II) and the Parkinson-related protein deglycase 1 (DJ-1), by recording fast 2D ^1H , ^{13}C NMR spectra of both intact cells and cell lysates. The unlabelled precursors, as well as the corresponding heavy isotope containing isotopologues (compounds **1***-**5***) are shown in scheme 1 and have been synthesized using previously reported protocols



Scheme 1 Amino acid precursors used in this study. p-Hydroxyphenylpyruvate **1**; Phenylpyruvate **2**; α -Ketoisocaproate **3**; α -Ketoisovalerate **4**; 4-Methylthio- α -ketobutanoate **5**; 3-Methyl- α -ketovalerate **6**; 4-(2-Carboxy-2-hydroxyvinyl)-*1H*-imidazolium chloride **7**; ([3,5-

$^{13}\text{C}_2$, 2,6- $^2\text{H}_2$] p-Hydroxyphenylpyruvate **1***; ([3,5- $^{13}\text{C}_2$, 2,4,6- $^2\text{H}_3$] Phenylpyruvate **2***; [5- ^{13}C , 3,3,4,5,5,5- $^2\text{H}_6$] α -Ketoisocaproate **3***; [4- ^{13}C , 3, 4,4,4- $^2\text{H}_4$] α -Ketoisovalerate **4***; 4-[^{13}C] Methylthio- α -ketobutanoate **5***

(Lichtenecker et al. 2004, 2013a; Fischer et al. 2007; Lichtenecker 2014).

Materials and methods

Plasmid design

Following the procedures outlined in previous studies, DNA sequences encoding for carbonic anhydrase 2 (CA II, NP_000058.1), deglycase protein 1 (DJ-1, NP_009193.2), and empty vector (pHL-empty) were cloned in the pHLsec vector, deprived of the secretion sequence (Aricescu et al. 2006).

Custom medium preparation

The composition of the custom-made medium was made accordingly to that of commercial high-glucose Dulbecco's modified Eagle's medium (DMEM, Sigma D6546), excluding the amino acid(s) to be labelled or to be replaced with precursors. The components were weighed on an analytical scale and dissolved in ultrapure water. The solution was left stirring overnight at 4 °C and the pH was brought to 7.4. The medium was then filtered under a laminar flow hood with a 0.22 µm filter and stored at -80 °C in 50 mL aliquots. The labelled amino acids or precursors were supplemented to the medium from concentrated stock solutions at the time of transfection.

Expression tests

HEK293T cells (ATCC CRL3216) were cultured in T25 flasks with DMEM (Life Technologies). At 90–95% confluence the cells were transiently transfected with the vectors in a custom-made medium where one amino acid was substituted with the corresponding precursor (MAG-LAB) in different concentrations with respect to the concentration of the amino acid (aa: precursor 1:1, 1:2, 1:5). The transfection was carried out following a previously reported protocol (Barbieri et al. 2016), in which vector and polyethylenimine (PEI, Sigma-Aldrich) were mixed in a 1:2 ratio (8.3 µg of DNA: 16.7 µg of PEI). After transfection the cells were incubated for 48 h at 37 °C, 5% CO₂, with custom-made DMEM supplemented with 2% (v/v) fetal bovine serum (FBS, Life Technologies) and 100 µg/mL penicillin–streptomycin (Life Technologies), respectively. It was previously shown that CA II and DJ-1 after 48 h of transient expression are correctly folded and do not undergo degradation (Barbieri et al. 2018; Luchinat et al. 2020a). The cells were then harvested in 150 µL of phosphate buffered saline (PBS, Life Technologies), lysed by freeze and thaw and the soluble part

of the lysate was collected after centrifugation. The expression level of CA II was evaluated by Coomassie-stained SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The loading sample, consisting in 7 µL of a solution made by diluting the soluble lysate 1:3 in 100mM Tris-HCl pH 6.8, 12.5% (v/v) glycerol, 2% (w/v) sodium dodecyl sulfate (SDS), 0.1% (w/v) bromophenol blue, and 50 mM dithiothreitol (DTT), was loaded onto pre-cast gels (Bio-Rad) and run at 200 V for 30 min. The gels were then stained with Coomassie Blue dye (ProBlue Safe Stain, Giotto Biotech).

Protein quantification

CA II concentration in the lysates from each condition was quantified by densitometric analysis on Coomassie-stained SDS-PAGE. Three dilutions of the lysates of the expression tests of CA II with two doses of each precursor, and with commercial DMEM were loaded onto a pre-cast gel as previously described and compared to a linear regression standard curve made of four samples of purified CA II at increasing concentrations (1, 2, 5, 10 µM). The gel was stained with Coomassie Blue dye, imaged by ChemiDoc XRS (Bio-Rad), and analyzed with Image J. Densitometry was used to quantify the protein expression based on the standard curve.

NMR sample preparation

HEK293T cells cultured in T75 flasks with DMEM were transiently transfected in custom-made medium with one amino acid substituted by a 2x dose of the cognate labelled precursor. Cell transfection was performed following the aforementioned protocol scaled up for T75 flasks (25 µg of DNA: 50 µg of PEI). Protein expression was carried out for 48 h in custom-made DMEM as above. Cells were harvested by trypsinization and suspended in 180 µL of NMR buffer (DMEM, 70 mM HEPES, and 20% (v/v) D₂O), transferred in a 3 mm Shigemi NMR tube, gently pelleted and inserted in the NMR instrument for analysis. At the end of the NMR experiment the cells were collected and lysed in PBS buffer by freeze and thaw as described before. The soluble part of the lysate was supplemented with 10% D₂O, transferred in a 3 mm NMR tube and inserted in the NMR instrument for analysis. For samples supplied with two precursors at the same time, the experiment was carried out with the same protocol using a custom-made medium containing both precursors and lacking the corresponding amino acids.

NMR spectra acquisition and analysis

All NMR spectra were recorded at 310 K on a 950 MHz ¹H Avance III (Bruker) with pulse sequences, acquisition

Table 1 NMR acquisition parameters

Precursor	1*	2*	3*	4*	5*	1*+2*	3*+4*
Pulse sequence	SOFAST HMQC	SOFAST HMQC	XL ALSO- FAST HMQC	ALSOFAST HMQC	XL ALSO- FAST HMQC	SOFAST HMQC	XL ALSO- FAST HMQC
Duration (minutes)	44	44	61	46	59	44	61
¹H spectral width (ppm)	17.5	17.5	13	16	13	17.5	13
¹³C spectral width (ppm)	13	13	30	25	20	16	30
¹³C frequency offset (ppm)	117.6	130.6	25.1	22.6	17.6	124.6	25.1
¹³C FID size							
in-cells	64	64	128	128	128	80	152
lysate	96	96	256	256	256	116	256
Dummy scans	64	64	64	64	64	64	64
Scans							
in-cells	96	96	64	64	64	80	56
lysate	64	64	32	32	32	56	32
Inter-scan delay (ms)	350	350	350	250	350	350	350

Table 2 Labelled precursors employed in this study

Precursor	Extended name	Labelling scheme
1*	[¹³ C ₂ , ² H ₂] Hydroxyphenylpyruvate	Aromatic ¹³ C- ¹ H, deuterated ring
2*	[¹³ C ₂ , ² H ₃] Phenylpyruvate	Aromatic ¹³ C- ¹ H, deuterated ring
3*	[¹³ C, ² H ₆] α-Ketoisocaproate	Methyl ¹³ CH ₃ , deuterated side chain
4*	[¹³ C, ² H ₄] α-Ketoisovalerate	Methyl ¹³ CH ₃ , deuterated side chain
5*	[¹³ C] Methylthio- α-ketobutanoate	Methyl ¹³ CH ₃

windows and parameters specifically selected for each labelled residue (Table 1). Aromatic-labelled samples were analyzed with 2D ¹H-¹³C SOFAST HMQC spectra (Schanda and Brutscher 2005; Sathyamoorthy et al. 2014), while methyl-labelled samples were analyzed with either 2D ¹H-¹³C ALSOFAST or 2D ¹H-¹³C XL-ALSOFAST HMQC spectra (Mueller 2008; Rößler et al. 2020). The experiments were performed with the same spectral window for both in-cell and lysate samples; for lysates we increased the size and decreased the number of scans per increment to increase the FID resolution while keeping constant the duration of the experiment (Table 1). Peak counting was performed automatically in Topspin and manually edited to account for partial peak overlap and to exclude peaks arising from the cellular background. CA II concentration in the cell lysates analyzed by NMR was estimated by comparing the signal intensity in the imino region of 1D ¹H NMR spectra, which is free from cellular background signals, with those of a reference cell lysate sample containing CA II recorded with identical acquisition parameters. The concentration of CA II in the reference lysate sample (240 μM) was determined by titration with the CA inhibitor methazolamide (MZA), previously shown to bind CA II in the slow exchange regime (Luchinat et al. 2020a).

CA II-ligand interaction

CA II was overexpressed in presence of compound 4* and treated with inhibitors to validate the labelling method for

the study of interactions. After transfection and incubation for 48 h the cells were treated with one of the following inhibitors: acetazolamide (AAZ) 50 μM for 2 h, MZA 10 μM for 1 h, ethoxzolamide (ETZ) 10 μM for 1 h. After treatment, cells were collected and the samples were prepared and analyzed as reported above.

Results

Protein expression with α-ketoacid precursors

Custom-made media were prepared, in which a single α-ketoacid precursor was supplemented in place of the corresponding amino acid. Tyrosine precursor 1, Phenylalanine precursor 2, Leucine precursor 3, Valine precursor 4, Methionine precursor 5, Isoleucine precursor 6 and Histidine precursor 7 were tested (Scheme 1). For subsequent NMR analysis, isotope-labelled precursors were employed (1*-5*, Scheme 1 and Table 2).

Protein expression efficiency in media supplemented with each precursor was tested by evaluating the expression level of CA II by SDS-PAGE. Each expression test was carried out with increasing amount of precursor (1x, 2x, and 5x doses relative to the concentration of the corresponding amino acid in the commercial DMEM) and compared to the expression in commercial DMEM (Fig. 2). For all precursors tested, the 2x dose was sufficient to achieve expression levels comparable to those obtained with the corresponding

amino acid. Lower levels were obtained with both 1x dose and 5x dose of precursor, the latter being most likely a result of cytotoxic effects due to the high concentration of precursor (Fig. 1). Quantitative analysis revealed that in the lysates from cells treated with DMEM (either commercial or custom-made) CA II concentration was $\sim 100 \mu\text{M}$, whereas with 1x, 2x and 5x doses of precursor CA II was $\sim 35\text{--}70 \mu\text{M}$, $70\text{--}100 \mu\text{M}$ and $45\text{--}85 \mu\text{M}$, respectively (Fig. S1).

In *E. coli*, α -ketoisovalerate **4**, acts as a precursor for both valine and leucine (Lichtenecker et al. 2013b). However, the pathway for the synthesis of leucine from α -ketoisovalerate is known to be inactive in mammalian cells (Neinast et al. 2019). Nevertheless, protein expression was tested in media

lacking only valine, only leucine, or both. As expected, expression was only achieved in the medium lacking valine, confirming that α -ketoisovalerate only works as a valine precursor in mammalian cells (Fig. 1B).

Incorporation of isotopically labelled precursors

For a subset of the above precursors (Table 2), protein side chain-specific isotopic labelling was evaluated through NMR spectroscopy in both intact cells and cell lysates. To this aim, two model globular proteins, CA II and DJ-1, were transiently overexpressed in presence of isotope-labelled precursors at the optimal doses reported above. To correctly

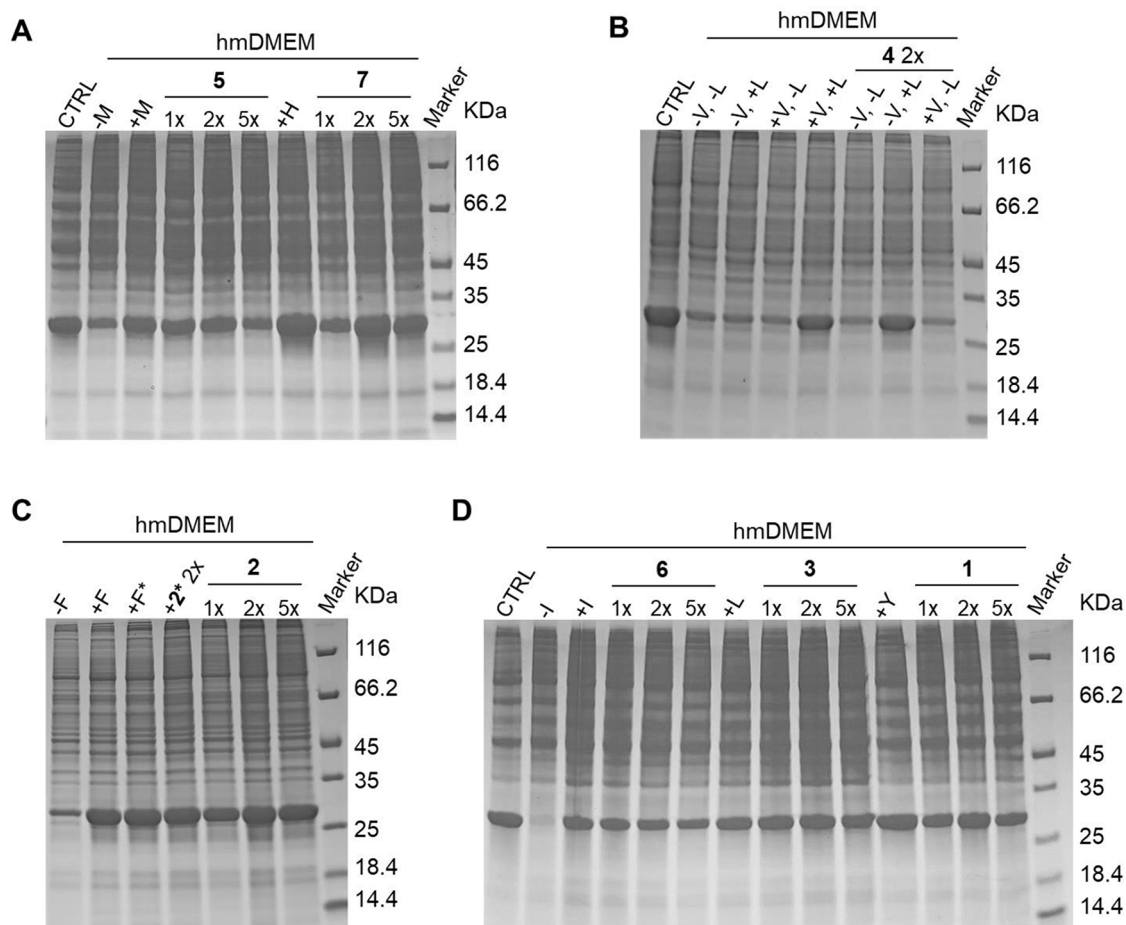


Fig. 1 Protein expression tests. **(A)** HEK293T overexpressing CAII, incubated with commercial DMEM, custom-made DMEM w/o Met, custom-made complete DMEM, custom-made DMEM w/ Methionine precursor **5** 1x, custom-made DMEM w/ **5** 2x, custom-made DMEM w/ **5** 5x, custom-made complete DMEM, custom-made DMEM w/ Histidine precursor **7** 1x, custom-made DMEM w/ **7** 2x, custom-made DMEM w/ **7** 5x, protein marker. **(B)** HEK293T overexpressing CAII, incubated with commercial DMEM, custom-made DMEM w/o Val and Leu, custom-made DMEM w/o Val, custom-made DMEM w/o Leu, custom-made complete DMEM, custom-made DMEM w/ Valine precursor **4** 2x w/o Val and Leu, custom-made DMEM w/ **4** 2x w/o Val, custom-made DMEM w/ **4** 2x w/o Leu, protein marker. **(C)** HEK293T

overexpressing CAII, incubated with custom-made DMEM w/o Phe, custom-made complete DMEM, custom-made DMEM w/ Phenylalanine precursor **2**, custom-made DMEM w/ labelled Phenylalanine precursor **2***, custom-made DMEM w/ **2** 1x, custom-made DMEM w/ **2** 2x, custom-made DMEM w/ **2** 5x, protein marker. **(D)** HEK293T overexpressing CAII, incubated with commercial DMEM, custom-made DMEM w/o Ile, custom-made complete DMEM, custom-made DMEM w/ Isoleucine precursor **6** 1x, custom-made DMEM w/ **6** 2x, custom-made DMEM w/ **6** 5x, custom-made complete DMEM, custom-made DMEM w/ Tyrosine precursor **1** 1x, custom-made DMEM w/ **1** 2x, custom-made DMEM w/ **1** 5x, protein marker

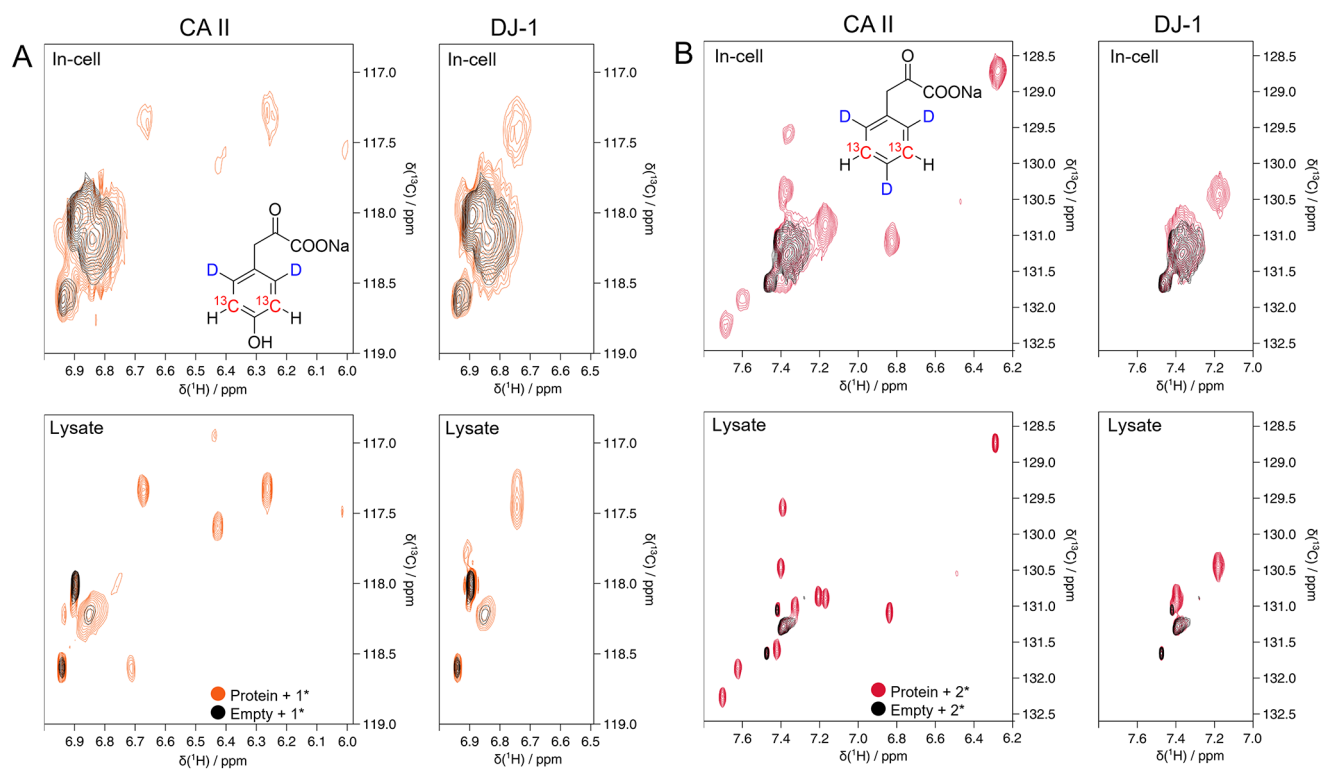


Fig. 2 NMR spectra of tyrosine and phenylalanine-labelled proteins. **(A)** Top: In-cell spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **1*** (orange); HEK293T cells transfected with empty vector, incubated with **1*** (black). Bottom: In-solution spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~40 μM , left) and DJ-1 (right), incubated with **1*** (orange); lysate of cells transfected with empty vector, incubated with **1*** (black). **(B)**

Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **2*** (red); HEK293T cells transfected with empty vector, incubated with **2*** (black). Bottom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~40 μM , left) and DJ-1 (right), incubated with **2*** (red); lysate of cells transfected with empty vector, incubated with **2*** (black)

Table 3 Observed vs. expected peaks detected in the NMR spectra of cell lysates. ^a Excluding N-terminal methionine

Precursor	Corresponding amino acid	Observed/expected peaks	
		CA II	DJ-1
1*	Tyr	5/8	3/3
2*	Phe	10/12	2/3
3*	Leu	50/52	35/36
4*	Val	34/34	36/38
5*	Met	1/1 ^a	4/4 ^a

identify the signals arising from the isotope-labelled over-expressed proteins, the spectra were compared with those obtained from control cells transfected with an empty vector and incubated in the same labelled media.

Aromatic labelling

We tested the incorporation of tyrosine and phenylalanine precursors **1*** and **2*** via in-cell and in-lysate NMR (Fig. 2) and compared the spectra of proteins obtained from a precursor-containing medium to those obtained from a medium containing the corresponding side-chain labelled

amino acids, which have been synthesized according to literature (Young et al. 2021) (Supp. Fig. S2). The ensuing in-cell NMR spectra revealed well-defined peaks for the side chain-labelled amino acids. The same peaks were better resolved in the lysate NMR spectra. Peak counting revealed a lower number of peaks than expected (Table 3). Notably, however, the same number of signals was also observed in the spectra obtained using the cognate amino acids (Supp. Fig. S2), indicating that the missing peaks do not arise from incomplete incorporation, but are likely caused by signal overlap and/or exchange broadening. Taken together, these results confirm that the α -ketoacids are transformed in the corresponding amino acids and subsequently integrated in the protein sequence at the correct positions.

Methyl labelling

We performed in-cell and in-lysate NMR experiments on protein samples methyl- ^{13}C labelled at valine, leucine, and methionine residues using compounds **3***, **4*** and **5***. The resulting in-cell NMR spectra showed well-resolved peaks corresponding to the side chain-labelled amino acids, which

were even better resolved in the lysate NMR spectra, confirming the incorporation in the protein sequence (Fig. 3). In the case of valine and leucine precursors, where the methyl groups are not stereospecifically labelled (resulting in a racemic mixture), the number of expected peaks corresponds to twice as many residues in the protein sequence, due to the labelling of both prochiral methyl groups (Fig. 3A, B). Despite the relatively low protein concentrations, ranging between 25 and 50 μM (Supp. Fig. S3), most of the expected methyl peaks were observed in the lysate spectra (Table 3), thanks to the high sensitivity and spectral resolution attained by selective methyl labelling. To assess the contribution of natural abundance ^{13}C in the spectra of selectively labelled samples, a lysate from cells expressing CA II in unlabelled medium was compared to a lysate containing leucine-labelled CA II. No signals from unlabelled protein were detected within the duration of the experiment, indicating that ^{13}C natural abundance does not interfere with the detection of methyl-labelled proteins (Supp. Fig. S4).

Combined precursor incorporation

We then combined multiple precursors in the same culture medium to assess whether the expression and incorporation levels remained unchanged. We chose valine and leucine precursors, which share the same transaminase enzyme, the branched-chain amino acid aminotransferase (BCAT) (Neinast et al. 2019; Toyokawa et al. 2021). This test would thus provide further information on enzymatic overload. Therefore, a custom culture medium containing Val and Leu precursor was made and the NMR experiments were performed overexpressing DJ-1. The resulting in-cell and lysate spectra show a good DJ-1 expression level, similar to the one of DJ-1 with the precursors used separately, and an efficient incorporation of both labelled amino acids (Fig. 4).

Application of the method

To show an example of a possible application of this labelling technique, we chose a set of well-characterized sulfonamide-derived CA inhibitors: Acetazolamide (AAZ) and methazolamide (MZA) which are currently in use as drugs to treat glaucoma, and ethoxzolamide (ETZ) which is an inhibitor of CAs in proximal renal tubules, widely used as a diuretic. All three compounds were previously shown to tightly bind CA II by in vitro and in-cell NMR (Luchinat et al. 2020a, b). We selected the valine precursor 4* as Val residues are present in the CA II ligand binding site. Indeed, from the spectra comparison it is possible to appreciate the change in the shifts of those valine residues involved in the interaction or experiencing an environment variation due to the proximity of the ligand (Fig. 5).

Discussion and conclusions

Exclusive isotope labelling of backbone positions is often not sufficient to unfold the full potential of NMR spectroscopy and perform in-depth protein characterization. Special isotope patterns in the side chains are required to investigate interaction surfaces via chemical shift perturbation or study side-chain dynamics in spin relaxation experiments. However, selective isotope labelling is often compromised by high costs for the heavy isotope-containing amino acids needed. Metabolic precursors such as α -ketoacids can be efficiently synthesized from low-cost heavy isotope-containing starting compounds; the corresponding protocols have been summarized in various reviewing articles (Schörghuber et al. 2018; Kerfah et al. 2015; Tugarinov et al. 2006; Rowlinson et al. 2022; Lichtenecker et al. 2015). Many of these precursor isotopologues are even commercially available today. Our results show that using precursors to introduce side chain-labelled amino acids is not limited to *E. coli* expression but can be transferred to mammalian cell hosts. This is especially useful when bacterial or yeast hosts are not viable options to produce a mammalian protein, which is usually the case for large proteins or proteins that undergo post-translational modifications. We furthermore demonstrated that the efficacy of the method is guaranteed even in the case of precursor combinations, granting an amino acid-selective labelling on more than one amino acid, thus expanding the possible applications of this labelling technique. The successful expression of proteins in the presence of the precursors employed in this work is allowed by the presence of transaminases in cultured mammalian cells: the reversible transamination of Ile, Leu, and Val is catalyzed by branched chain amino acid amino transferase (BCAT), Tyr and Phe are substrates of tyrosine aminotransferase (TAT) and glutamic-oxaloacetic transaminase (GOT), the transamination of Met is catalyzed by glutamine transaminase K (GTK), and His is a substrate of both glutamine transaminase L (GTL) and serine-pyruvate transaminase (SPT) (Caligiore et al. 2022; Neinast et al. 2019; Toyokawa et al. 2021; Sivaraman and Kirsch 2006; Mehre et al. 2010; Cooper 2004). Concerning the general applicability of the method, our results suggest that α -ketoacid precursors of other amino acids could be employed, provided that they are recognized by transaminases expressed in the cell culture. GTK and TAT are reported to catalyze the transamination of Trp with high efficiency, whereas a similar activity for Lys and Arg is not reported (Caligiore et al. 2022).

Furthermore, side chain isotope labelling may not be straightforward for some amino acids (Ala, Asp, Asn, Pro, Glu) that are directly connected to core cellular metabolic pathways, such as glycolysis and the TCA cycle. In such cases, adding the corresponding α -ketoacid precursors to

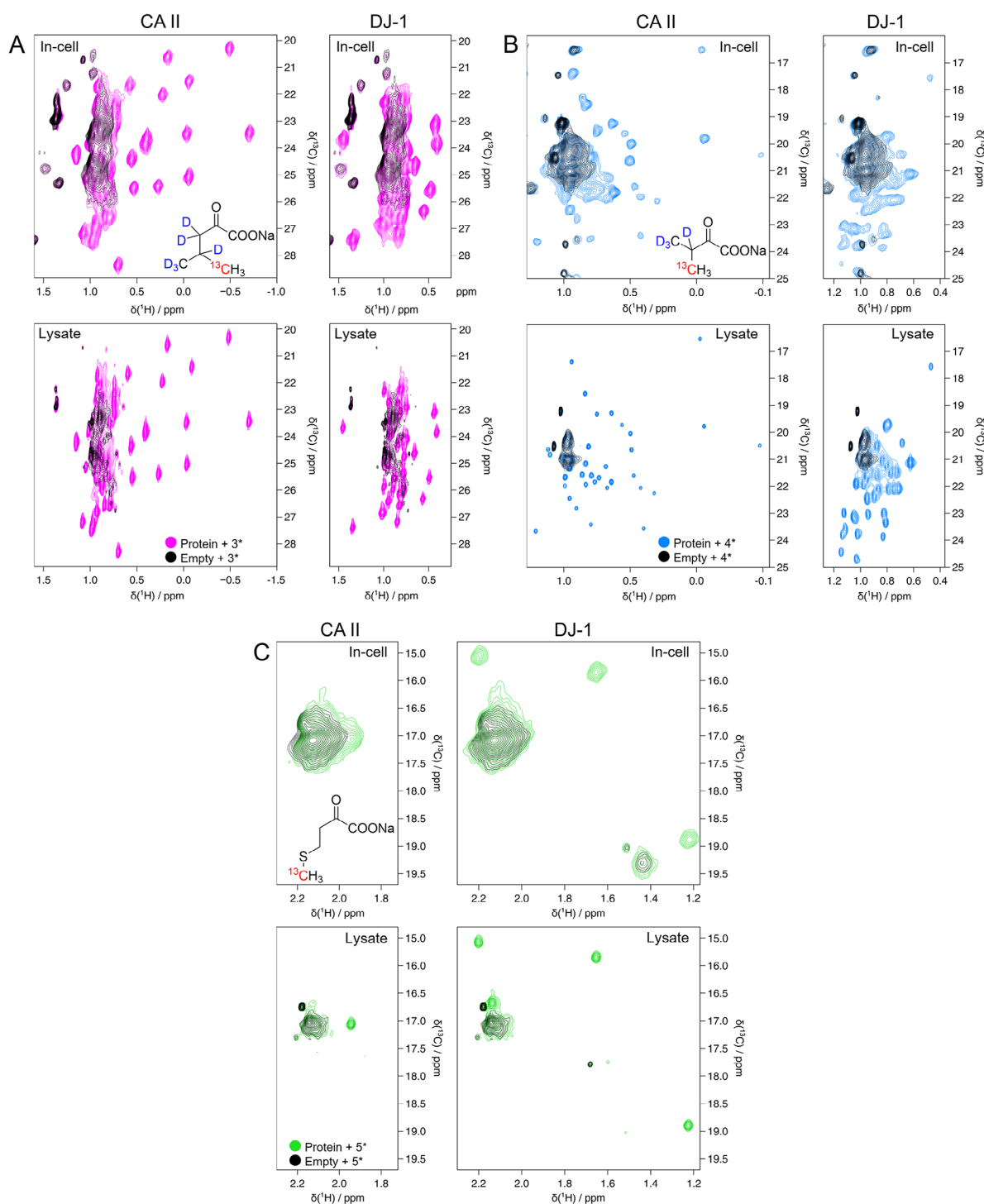


Fig. 3 NMR spectra of valine, leucine and methionine-labelled proteins. **(A)** Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **3*** (magenta); HEK293T cells overexpressing empty vector, incubated with **3*** (black). Bottom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~50 μ M, left) and DJ-1 (right), incubated with **3*** (magenta); lysate of cells transfected with empty vector, incubated with **3*** (black). **(B)** Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **4*** (light blue); HEK293T cells transfected with empty vector, incubated with **4*** (black). Bot-

tom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~25 μ M, left) and DJ-1 (right), incubated with **4*** (light blue); lysate of cells transfected with empty vector, incubated with **4*** (black). **(C)** Top: In-cell NMR spectra on HEK293T cells overexpressing CA II (left) and DJ-1 (right), incubated with **5*** (green); HEK293T cells transfected with empty vector, incubated with **5*** (black). Bottom: NMR spectra on HEK293T lysate. Lysate of cells overexpressing CA II (~25 μ M, left) and DJ-1 (right), incubated with **5*** (green); lysate of cells transfected with empty vector, incubated with **5*** (black)

Fig. 4 Simultaneous labelling with valine and leucine. Left: In-cell spectra on HEK293T overexpressing DJ-1, incubated with **3*** and **4*** (magenta); HEK293T cells overexpressing empty vector, incubated with **3*** and **4*** (black). Right: In-solution spectra on HEK293T lysate overexpressing DJ-1, incubated with **3*** and **4*** (magenta); lysate of cells overexpressing empty vector, incubated with **3*** and **4*** (black)

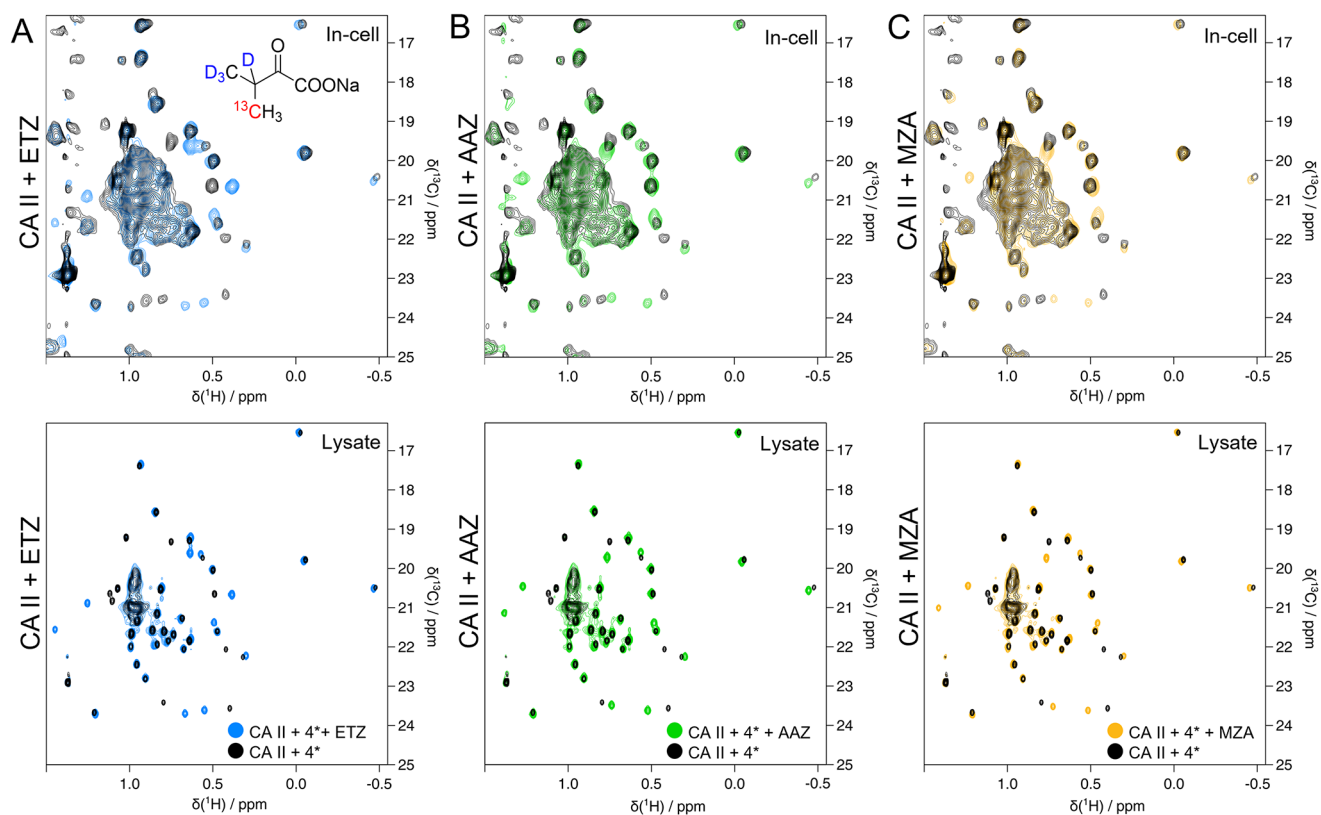
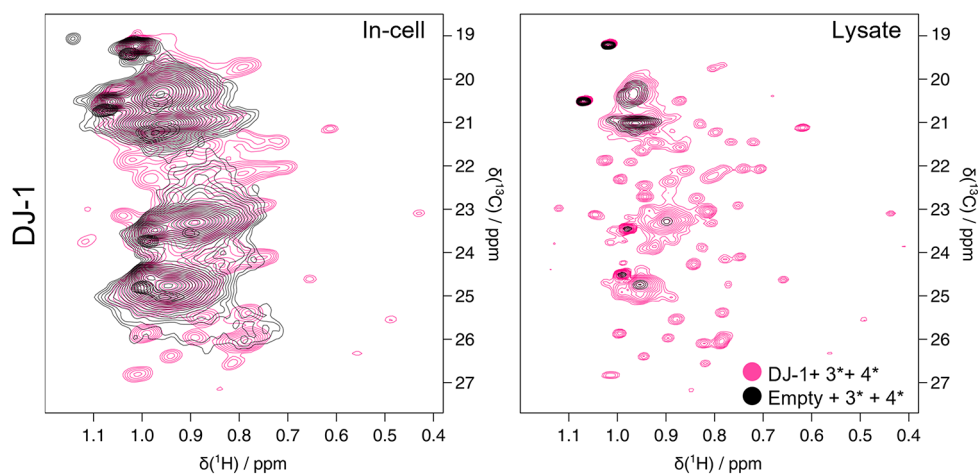


Fig. 5 Ligand binding to CA II. (A) Top: In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** and treated with ETZ (light blue); In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** (black) Bottom: NMR spectra on HEK293T lysate overexpressing CA II, incubated with **4*** and treated with ETZ (light blue); In-solution spectra on HEK293T lysate overexpressing CA II, incubated with **4*** (black). (B) Top: In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** and treated with AAZ (green); In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** (black). Bottom: NMR spectra on HEK293T

lysate overexpressing CA II, incubated with **4*** and treated with AAZ (green); In-solution spectra on HEK293T lysate overexpressing CA II, incubated with **4*** (black). (C) Top: In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** and treated with MZA (yellow); In-cell NMR spectra on HEK 293T overexpressing CA II, incubated with **4*** (black). Bottom: NMR spectra on HEK293T lysate overexpressing CA II, incubated with **4*** and treated with MZA (yellow); In-solution spectra on HEK293T lysate overexpressing CA II, incubated with **4*** (black)

the culture medium will likely result in the unwanted labeling of other amino acids in the expressed proteins. Notably, however, the above amino acids are not predominant in protein-ligand and protein-protein interfaces, and therefore their labeling would not be the first choice for studying such phenomena.

As a first example, we demonstrated how this labelling technique results in well-resolved methyl resonances in the NMR spectra of two globular proteins, the 29 kDa CA II and the 40 kDa DJ-1 homodimer. Addition of selected known ligands led to substantial perturbation of the CA II chemical shifts, which could be recorded in-lysate, as well as in-cell. This observation implies that combinations of precursors and mammalian expression systems are generally applicable to investigate protein-ligand- and, in principle, protein-protein interactions. Overall, we believe that introducing α -ketoacid precursors to mammalian cell-based protein expression will greatly facilitate the application of NMR to study structure and function of challenging human proteins in a purified state, in cell lysates, or even in intact cells.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10858-024-00447-6>.

Author contributions R.K., E.L. and Lu.B. conceived the work; M.R., Le.B. R.J.L., R.K., E.L. and Lu.B. designed the experiments; M.H., S.K. and R.J.L. synthesized the precursor compounds; M.R. and Le.B. prepared the NMR samples; M.R. and E.L. collected, processed and analyzed the NMR data; M.R. and E.L. drafted the manuscript; M.R., Le.B., R.J.L., R.K., E.L. and Lu.B. contributed to the manuscript; Lu.B. secured funding.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement. This work was supported by Instruct-ERIC, a Landmark ESFRI project, and specifically by the CERM/CIRMMP Italian Instruct Centre, and by Fragment-Screen, grant agreement No. 101094131, funded by the Horizon Europe program of the European Commission. The authors acknowledge the support by the Italian Ministry for University and Research (MUR) FOE funding to CERM/CIRMMP. E.L. contract is funded by MUR through the Dipartimento di Eccellenza 2023–2027 (DICUS 2.0) funding to the Department of Chemistry “Ugo Schiff” of the University of Florence. Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement.

Data availability The raw NMR data will be made available upon reasonable request.

Declarations

Competing interests R.J.L. and R.K. are shareholders in the company Mag-Lab, which is mentioned in the affiliations. R.J.L., M.H. and S.K. are employed at Mag-Lab.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the

source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aricescu AR, Lu W, Jones EY (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr* 62:1243–1250. <https://doi.org/10.1107/S0907444906029799>
- Banci L, Barbieri L, Bertini I et al (2013) Atomic-resolution monitoring of protein maturation in live human cells by NMR. *Nat Chem Biol* 9:297–299. <https://doi.org/10.1038/nchembio.1202>
- Barbieri L, Luchinat E, Banci L (2015) Protein interaction patterns in different cellular environments are revealed by in-cell NMR. *Sci Rep* 5:14456. <https://doi.org/10.1038/srep14456>
- Barbieri L, Luchinat E, Banci L (2016) Characterization of proteins by in-cell NMR spectroscopy in cultured mammalian cells. *Nat Protoc* 11:1101–1111. <https://doi.org/10.1038/nprot.2016.061>
- Barbieri L, Luchinat E, Banci L (2018) Intracellular metal binding and redox behavior of human DJ-1. *J Biol Inorg Chem JBIC Publ Soc Biol Inorg Chem* 23:61–69. <https://doi.org/10.1007/s00775-017-1509-5>
- Burz DS, Breindel L, Shekhtman A (2019) Improved sensitivity and resolution of in-cell NMR spectra. *Methods Enzymol* 621:305–328. <https://doi.org/10.1016/bs.mie.2019.02.029>
- Caligiore F, Zangelmi E, Vetro C et al (2022) Human cytosolic transaminases: side activities and patterns of discrimination towards physiologically available alternative substrates. *Cell Mol Life Sci* 79:421. <https://doi.org/10.1007/s00018-022-04439-3>
- Cooper AJL (2004) The role of glutamine transaminase K (GTK) in sulfur and α -keto acid metabolism in the brain, and in the possible bioactivation of neurotoxicants. *Neurochem Int* 44:557–577. <https://doi.org/10.1016/j.neuint.2003.12.002>
- Fischer M, Kloiber K, Häusler J et al (2007) Synthesis of a 13 C-Methyl-group-labeled methionine precursor as a useful Tool for simplifying protein structural analysis by NMR spectroscopy. *ChemBioChem* 8:610–612. <https://doi.org/10.1002/cbic.200600551>
- Goto NK, Gardner KH, Mueller GA et al (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ 1) methyl-protonated 15 N-, 13 C-, 2H-labeled proteins. *J Biomol NMR* 13:369–374. <https://doi.org/10.1023/a:1008393201236>
- Kerfah R, Plevin MJ, Sounier R et al (2015) Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr Opin Struct Biol* 32:113–122. <https://doi.org/10.1016/j.sbi.2015.03.009>
- Lichtenecker RJ (2014) Synthesis of aromatic 13 C/2H- α -ketoacid precursors to be used in selective phenylalanine and tyrosine protein labelling. *Org Biomol Chem* 12:7551–7560. <https://doi.org/10.1039/C4OB01129E>
- Lichtenecker R, Ludwiczek ML, Schmid W, Konrat R (2004) Simplification of protein NOESY Spectra using Bioorganic Precursor synthesis and NMR spectral editing. *J Am Chem Soc* 126:5348–5349. <https://doi.org/10.1021/ja049679n>
- Lichtenecker RJ, Coudeville N, Konrat R, Schmid W (2013a) Selective isotope labelling of leucine residues by using α -Ketoacid

- precursor compounds. *ChemBioChem* 14:818–821. <https://doi.org/10.1002/cbic.201200737>
- Lichtenecker RJ, Weinhäupl K, Reuther L et al (2013b) Independent valine and leucine isotope labeling in *Escherichia coli* protein overexpression systems. *J Biomol NMR* 57:205–209. <https://doi.org/10.1007/s10858-013-9786-y>
- Lichtenecker RJ, Weinhäupl K, Schmid W, Konrat R (2013c) α -Ketoacids as precursors for phenylalanine and tyrosine labeling in cell-based protein overexpression. *J Biomol NMR* 57:327–331. <https://doi.org/10.1007/s10858-013-9796-9>
- Lichtenecker RJ, Schörghuber J, Bisaccia M (2015) Synthesis of metabolic amino acid precursors: tools for selective isotope labeling in cell-based protein overexpression. *Synlett* 26:2611–2616. <https://doi.org/10.1055/s-0034-1378878>
- Luchinat E, Barbieri L, Cremonini M et al (2020a) Drug screening in human cells by NMR spectroscopy allows the Early Assessment of Drug Potency. *Angew Chem Int Ed Engl* 59:6535–6539. <https://doi.org/10.1002/anie.201913436>
- Luchinat E, Barbieri L, Cremonini M et al (2020b) Intracellular Binding/Unbinding Kinetics of approved drugs to Carbonic anhydrase II observed by in-cell NMR. *ACS Chem Biol* 15:2792–2800. <https://doi.org/10.1021/acscchembio.0c00590>
- Mehere P, Han Q, Lemkul JA et al (2010) Tyrosine aminotransferase: biochemical and structural properties and molecular dynamics simulations. *Protein Cell* 1:1023–1032. <https://doi.org/10.1007/s13238-010-0128-5>
- Mueller L (2008) Alternate HMQC experiments for recording HN and HC-correlation spectra in proteins at high throughput. *J Biomol NMR* 42:129–137. <https://doi.org/10.1007/s10858-008-9270-2>
- Neinast M, Murashige D, Arany Z (2019) Branched chain amino acids. *Annu Rev Physiol* 81:139–164. <https://doi.org/10.1146/annurev-physiol-020518-114455>
- Ninković DB, Blagojević Filipović JP, Hall MB et al (2020) What is special about aromatic–aromatic interactions? Significant attraction at large horizontal displacement. *ACS Cent Sci* 6:420–425. <https://doi.org/10.1021/acscentsci.0c00005>
- Rößler P, Mathieu D, Gossert AD (2020) Enabling NMR studies of high Molecular Weight systems without the need for Deuteration: the XL-ALSOFASST experiment with delayed decoupling. *Angew Chem Int Ed* 59:19329–19337. <https://doi.org/10.1002/anie.202007715>
- Rowlinson B, Crublet E, Kerfah R, Plevin MJ (2022) Specific isotopic labelling and reverse labelling for protein NMR spectroscopy: using metabolic precursors in sample preparation. *Biochem Soc Trans* 50:1555–1567. <https://doi.org/10.1042/BST20210586>
- Sathyamoorthy B, Lee J, Kimsey I et al (2014) Development and application of aromatic [(13)C, (1)H] SOFAST-HMQC NMR experiment for nucleic acids. *J Biomol NMR* 60:77–83. <https://doi.org/10.1007/s10858-014-9856-9>
- Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J Am Chem Soc* 127:8014–8015. <https://doi.org/10.1021/ja051306e>
- Schörghuber J, Geist L, Platzer G et al (2018) Late metabolic precursors for selective aromatic residue labeling. *J Biomol NMR* 71:129–140. <https://doi.org/10.1007/s10858-018-0188-z>
- Serrano L, Bycroft M, Fersht AR (1991) Aromatic-aromatic interactions and protein stability. *J Mol Biol* 218:465–475. [https://doi.org/10.1016/0022-2836\(91\)90725-L](https://doi.org/10.1016/0022-2836(91)90725-L)
- Sivaraman S, Kirsch JF (2006) The narrow substrate specificity of human tyrosine aminotransferase – the enzyme deficient in tyrosinemia type II. *FEBS J* 273:1920–1929. <https://doi.org/10.1111/j.1742-4658.2006.05202.x>
- Theillet F-X, Luchinat E (2022) In-cell NMR: why and how? *Prog Nucl Magn Reson Spectrosc.* 132–133:1–112. <https://doi.org/10.1016/j.pnmrs.2022.04.002>
- Toyokawa Y, Koonthongkaew J, Takagi H (2021) An overview of branched-chain amino acid aminotransferases: functional differences between mitochondrial and cytosolic isozymes in yeast and human. *Appl Microbiol Biotechnol* 105:8059–8072. <https://doi.org/10.1007/s00253-021-11612-4>
- Tugarinov V, Kanelis V, Kay LE (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* 1:749–754. <https://doi.org/10.1038/nprot.2006.101>
- Werner K, Richter C, Klein-Seetharaman J, Schwalbe H (2008) Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy. *J Biomol NMR* 40:49–53. <https://doi.org/10.1007/s10858-007-9205-3>
- Young BM, Rossi P, Slavish PJ et al (2021) Synthesis of isotopically labeled, spin-isolated tyrosine and phenylalanine for protein NMR applications. *Org Lett* 23:6288–6292. <https://doi.org/10.1021/acs.orglett.1c02084>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.